

23369  
PT 1.2057 Pt  
E.M.: 1007. idet.  
(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number  
WO 01/00843 A2

(51) International Patent Classification: C12N 15/31,  
15/61, 1/21, 9/90, C07K 14/34, C12P 13/08, C12Q 1/68

(21) International Application Number: PCT/IB00/00923

(22) International Filing Date: 23 June 2000 (23.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/141,031 25 June 1999 (25.06.1999) US  
199 30 476.9 1 July 1999 (01.07.1999) DE  
60/142,101 2 July 1999 (02.07.1999) US  
199 31 415.2 8 July 1999 (08.07.1999) DE  
199 31 418.7 8 July 1999 (08.07.1999) DE  
199 31 419.5 8 July 1999 (08.07.1999) DE  
199 31 420.9 8 July 1999 (08.07.1999) DE  
199 31 424.1 8 July 1999 (08.07.1999) DE  
199 31 428.4 8 July 1999 (08.07.1999) DE  
199 31 434.9 8 July 1999 (08.07.1999) DE  
199 31 435.7 8 July 1999 (08.07.1999) DE  
199 31 443.8 8 July 1999 (08.07.1999) DE  
199 31 453.5 8 July 1999 (08.07.1999) DE  
199 31 457.8 8 July 1999 (08.07.1999) DE  
199 31 465.9 8 July 1999 (08.07.1999) DE  
199 31 478.0 8 July 1999 (08.07.1999) DE  
199 31 510.8 8 July 1999 (08.07.1999) DE  
199 31 541.8 8 July 1999 (08.07.1999) DE  
199 31 573.6 8 July 1999 (08.07.1999) DE  
199 31 592.2 8 July 1999 (08.07.1999) DE  
199 31 632.5 8 July 1999 (08.07.1999) DE  
199 31 634.1 8 July 1999 (08.07.1999) DE  
199 31 636.8 8 July 1999 (08.07.1999) DE  
199 32 125.6 9 July 1999 (09.07.1999) DE  
199 32 126.4 9 July 1999 (09.07.1999) DE  
199 32 130.2 9 July 1999 (09.07.1999) DE  
199 32 186.8 9 July 1999 (09.07.1999) DE  
199 32 206.6 9 July 1999 (09.07.1999) DE  
199 32 227.9 9 July 1999 (09.07.1999) DE  
199 32 228.7 9 July 1999 (09.07.1999) DE  
199 32 229.5 9 July 1999 (09.07.1999) DE  
199 32 230.9 9 July 1999 (09.07.1999) DE  
199 32 922.2 14 July 1999 (14.07.1999) DE  
199 32 926.5 14 July 1999 (14.07.1999) DE  
199 32 928.1 14 July 1999 (14.07.1999) DE

199 33 004.2 14 July 1999 (14.07.1999) DE  
199 33 005.0 14 July 1999 (14.07.1999) DE  
199 33 006.9 14 July 1999 (14.07.1999) DE  
60/148,613 12 August 1999 (12.08.1999) US  
199 40 764.9 27 August 1999 (27.08.1999) DE  
199 40 765.7 27 August 1999 (27.08.1999) DE  
199 40 766.5 27 August 1999 (27.08.1999) DE  
199 40 832.7 27 August 1999 (27.08.1999) DE  
199 41 378.9 31 August 1999 (31.08.1999) DE  
199 41 379.7 31 August 1999 (31.08.1999) DE  
199 41 394.0 31 August 1999 (31.08.1999) DE  
199 41 396.7 31 August 1999 (31.08.1999) DE  
199 41 380.0 31 August 1999 (31.08.1999) DE  
199 42 077.7 3 September 1999 (03.09.1999) DE  
199 42 129.3 3 September 1999 (03.09.1999) DE  
199 42 076.9 3 September 1999 (03.09.1999) DE  
199 42 079.3 3 September 1999 (03.09.1999) DE  
199 42 086.6 3 September 1999 (03.09.1999) DE  
199 42 087.4 3 September 1999 (03.09.1999) DE  
199 42 088.2 3 September 1999 (03.09.1999) DE  
199 42 095.5 3 September 1999 (03.09.1999) DE  
199 42 124.2 3 September 1999 (03.09.1999) DE  
60/187,970 9 March 2000 (09.03.2000) US

(71) Applicant: BASF AKTIENGESELLSCHAFT  
[DE/DE]; D-67056 Ludwigshafen (DE).

(72) Inventors: POMPEJUS, Markus; Wenjenstr. 21,  
D-67251 Freinsheim (DE). KRÖGER, Burkhard; Im  
Waldhof 1, D-67117 Limburgerhof (DE). SCHRÖDER,  
Hartwig; Goethestr. 5, D-69226 Nussloch (DE).  
ZELDER, Oskar; Rossmarktstr. 27, D-67346 Speyer  
(DE). HABERHAUER, Gregor; Moselstr. 42, D-67117  
Limburgerhof (DE).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING METABOLIC PATHWAY PROTEINS

(57) Abstract: Isolated nucleic acid molecules, designated MP nucleic acid molecules, which encode novel MP proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing MP nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated MP proteins, mutated MP proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of MP genes in this organism.

WO 01/00843 A2

WO 01/00843 A2



IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

- *Without international search report and to be republished upon receipt of that report.*

been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

## 5 Summary of the Invention

The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be used to produce fine chemicals, the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related  
10 bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as metabolic pathway (MP) proteins.

*C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the  
15 degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The MP nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, *e.g.*, by fermentation processes. Modulation of the expression of the MP nucleic acids of the invention, or modification of the sequence of the MP nucleic acid molecules of the  
20 invention, can be used to modulate the production of one or more fine chemicals from a microorganism (*e.g.*, to improve the yield or production of one or more fine chemicals from a *Corynebacterium* or *Brevibacterium* species).

The MP nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the  
25 presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain  
30 whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to species pathogenic in humans, such as *Corynebacterium*

1 *diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

→ The MP nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms.

- 5 Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered *Corynebacterium* or *Brevibacterium* species.

The MP proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing an enzymatic step involved in the metabolism of certain fine chemicals, including amino acids, vitamins, cofactors, nutraceuticals,

- 2  
10 nucleotides, nucleosides, and trehalose. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey *et al.*, U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (*e.g.*, *lactofermentum*) (Yoshihama *et al.*, *J. Bacteriol.* 162: 591-597 (1985); Katsumata *et al.*, *J. Bacteriol.* 159: 306-311 (1984); and  
15 Santamaria *et al.*, *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals.

- This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an  
20 indirect effect of such manipulation. Specifically, alterations in *C. glutamicum* metabolic pathways for amino acids, vitamins, cofactors, nucleotides, and trehalose may have a direct impact on the overall production of one or more of these desired compounds from this organism. For example, optimizing the activity of a lysine biosynthetic pathway protein or decreasing the activity of a lysine degradative pathway  
25 protein may result in an increase in the yield or efficiency of production of lysine from such an engineered organism. Alterations in the proteins involved in these metabolic pathways may also have an indirect impact on the production or efficiency of production of a desired fine chemical. For example, a reaction which is in competition for an intermediate necessary for the production of a desired molecule may be eliminated, or a  
3  
30 pathway necessary for the production of a particular intermediate for a desired compound may be optimized. Further, modulations in the biosynthesis or degradation of, for example, an amino acid, a vitamin, or a nucleotide may increase the overall



ability of the microorganism to rapidly grow and divide, thus increasing the number and/or production capacities of the microorganism in culture and thereby increasing the possible yield of the desired fine chemical.

The nucleic acid and protein molecules of the invention may be utilized to  
5 directly improve the production or efficiency of production of one or more desired fine chemicals from *Corynebacterium glutamicum*. Using recombinant genetic techniques well known in the art, one or more of the biosynthetic or degradative enzymes of the invention for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose may be manipulated such that its function is modulated. For example, a  
10 biosynthetic enzyme may be improved in efficiency, or its allosteric control region destroyed such that feedback inhibition of production of the compound is prevented. Similarly, a degradative enzyme may be deleted or modified by substitution, deletion, or addition such that its degradative activity is lessened for the desired compound without impairing the viability of the cell. In each case, the overall yield or rate of production of  
15 the desired fine chemical may be increased.

It is also possible that such alterations in the protein and nucleotide molecules of the invention may improve the production of other fine chemicals besides the amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose through indirect mechanisms. Metabolism of any one compound is necessarily  
20 intertwined with other biosynthetic and degradative pathways within the cell, and necessary cofactors, intermediates, or substrates in one pathway are likely supplied or limited by another such pathway. Therefore, by modulating the activity of one or more of the proteins of the invention, the production or efficiency of activity of another fine chemical biosynthetic or degradative pathway may be impacted. For example, amino  
25 acids serve as the structural units of all proteins, yet may be present intracellularly in levels which are limiting for protein synthesis; therefore, by increasing the efficiency of production or the yields of one or more amino acids within the cell, proteins, such as biosynthetic or degradative proteins, may be more readily synthesized. Likewise, an alteration in a metabolic pathway enzyme such that a particular side reaction becomes  
30 more or less favored may result in the over- or under-production of one or more compounds which are utilized as intermediates or substrates for the production of a desired fine chemical.

This invention provides novel nucleic acid molecules which encode proteins, referred to herein as metabolic pathway proteins (MP), which are capable of, for example, performing an enzymatic step involved in the metabolism of molecules important for the normal functioning of cells, such as amino acids, vitamins, cofactors, nucleotides and nucleosides, or trehalose. Nucleic acid molecules encoding an MP protein are referred to herein as MP nucleic acid molecules. In a preferred embodiment, the MP protein performs an enzymatic step related to the metabolism of one or more of the following: amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (*e.g.*, cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an MP protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of MP-encoding nucleic acid (*e.g.*, DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth as the odd-numbered SEQ ID NOs in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7....), or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth as an odd-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7....), or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth as an even-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8....). The preferred MP proteins of the present invention also preferably possess at least one of the MP activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a

sequence having an even-numbered SEQ ID NO: in the Sequence Listing), *e.g.*, sufficiently homologous to an amino acid sequence of the invention such that the protein or portion thereof maintains an MP activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to perform an enzymatic reaction in a amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of the invention (*e.g.*, an entire amino acid sequence selected from those having an even-numbered SEQ ID NO in the Sequence Listing). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of the invention (encoded by an open reading frame shown in the corresponding odd-numbered SEQ ID NOs in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7....)).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (*e.g.*, an MP fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of the invention (*e.g.*, a sequence of one of the even-numbered SEQ ID NOs in the Sequence Listing) and is able to catalyze a reaction in a metabolic pathway for an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose, or one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO in the Sequence Listing). Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* MP protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, *e.g.*, recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an MP protein by culturing the host cell in a suitable medium. The MP protein  
5 can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an MP gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated MP sequence as a  
10 transgene. In another embodiment, an endogenous MP gene within the genome of the microorganism has been altered, *e.g.*, functionally disrupted, by homologous recombination with an altered MP gene. In another embodiment, an endogenous or introduced MP gene in a microorganism has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional MP protein. In still  
15 another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an MP gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the MP gene is modulated. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being  
20 particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

In another aspect, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of  
25 one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in the Sequence Listing as SEQ ID NOs 1 through 1156) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

Still another aspect of the invention pertains to an isolated MP protein or a  
30 portion, *e.g.*, a biologically active portion, thereof. In a preferred embodiment, the isolated MP protein or portion thereof can catalyze an enzymatic reaction involved in one or more pathways for the metabolism of an amino acid, a vitamin, a cofactor, a

nutraceutical, a nucleotide, a nucleoside, or trehalose. In another preferred embodiment, the isolated MP protein or portion thereof is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: in the Sequence Listing) such that the protein or portion thereof maintains the ability to catalyze an enzymatic reaction involved in one or more pathways for the metabolism of an amino acid, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose.

The invention also provides an isolated preparation of an MP protein. In preferred embodiments, the MP protein comprises an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing). In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) (encoded by an open reading frame set forth in a corresponding odd-numbered SEQ ID NO: of the Sequence Listing). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing). In other embodiments, the isolated MP protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) and is able to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated MP protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to a nucleotide sequence of one of the even-numbered SEQ ID NOs set forth in the Sequence Listing. It is also preferred that the preferred forms of MP proteins also have one or more of the MP bioactivities described herein.

The MP polypeptide, or a biologically active portion thereof, can be operatively linked to a non-MP polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the MP protein alone. In other preferred embodiments, this fusion protein, when introduced into a *C. glutamicum* pathway for the metabolism of an amino acid, vitamin, cofactor, nutraceutical, results in increased yields and/or efficiency of production of a desired fine chemical from *C. glutamicum*. In particularly preferred embodiments, integration of this fusion protein into an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway of a host cell modulates production of a desired compound from the cell.

In another aspect, the invention provides methods for screening molecules which modulate the activity of an MP protein, either by interacting with the protein itself or a substrate or binding partner of the MP protein, or by modulating the transcription or translation of an MP nucleic acid molecule of the invention.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an MP nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an MP nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates MP protein activity or MP nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates MP protein activity can be an agent which stimulates MP protein activity or MP nucleic acid expression.

Examples of agents which stimulate MP protein activity or MP nucleic acid expression include small molecules, active MP proteins, and nucleic acids encoding MP proteins that have been introduced into the cell. Examples of agents which inhibit MP activity or expression include small molecules, and antisense MP nucleic acid molecules.

5        Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant MP gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the  
10 introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

15

#### **Detailed Description of the Invention**

The present invention provides MP nucleic acid and protein molecules which are involved in the metabolism of certain fine chemicals in *Corynebacterium glutamicum*, including amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and  
20 trehalose. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (*e.g.*, where modulation of the activity of a lysine biosynthesis protein has a direct impact on the production or efficiency of production of lysine from that organism), or may have an indirect impact which nonetheless results in an increase of  
25 yield or efficiency of production of the desired compound (*e.g.*, where modulation of the activity of a nucleotide biosynthesis protein has an impact on the production of an organic acid or a fatty acid from the bacterium, perhaps due to improved growth or an increased supply of necessary co-factors, energy compounds, or precursor molecules). Aspects of the invention are further explicated below.

30

## I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include

5 organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described *e.g.* in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty

10 acids (*e.g.*, arachidonic acid), diols (*e.g.*, propane diol, and butane diol), carbohydrates (*e.g.*, hyaluronic acid and trehalose), aromatic compounds (*e.g.*, aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids,

15 Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane *et al.* (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN:

20 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

### *A. Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are

25 essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ullmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH:

30 Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids



have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3<sup>rd</sup> edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino acids – technical production and use, p. 466-502 in Rehm *et al.* (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E. (1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of  $\alpha$ -ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-

step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain  $\beta$ -carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3<sup>rd</sup> ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3<sup>rd</sup> ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

### 30 *B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses*

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although

they are readily synthesized by other organisms, such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (*e.g.*, polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B<sub>1</sub>) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B<sub>2</sub>) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B<sub>6</sub>' (*e.g.*, pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic

acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)- $\beta$ -alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of  $\beta$ -alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to  $\beta$ -alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B<sub>5</sub>), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the  $\alpha$ -ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which in turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B<sub>12</sub>) and porphyrines belong to a group of chemicals characterized by a tetrapyrrole ring system. The biosynthesis of vitamin B<sub>12</sub> is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B<sub>6</sub>, pantothenate, and

biotin. Only Vitamin B<sub>12</sub> is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

5 C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (*i.e.*, AMP) or as coenzymes (*i.e.*, FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (*e.g.* Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (*e.g.*, thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (*e.g.*, ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (*e.g.*, IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and

Related Compounds in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

- 5           The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology,
- 10   Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or
- 15   adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP).
- 20   The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

#### 25   D. Trehalose Metabolism and Uses

- Trehalose consists of two glucose molecules, bound in  $\alpha$ ,  $\alpha$ -1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto *et al.*, (1998) U.S.
- 30   Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) *J. Japan* 172: 97-102). Trehalose is produced by enzymes from

many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

## II. Elements and Methods of the Invention

5           The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as MP nucleic acid and protein molecules, which play a role in or function in one or more cellular metabolic pathways. In one embodiment, the MP molecules catalyze an enzymatic reaction involving one or more amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic  
10 pathways. In a preferred embodiment, the activity of the MP molecules of the present invention in one or more *C. glutamicum* metabolic pathways for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides or trehalose has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the MP molecules of the invention are modulated in activity, such that the  
15 *C. glutamicum* metabolic pathways in which the MP proteins of the invention are involved are modulated in efficiency or output, which either directly or indirectly modulates the production or efficiency of production of a desired fine chemical by *C. glutamicum*.

          The language, "MP protein" or "MP polypeptide" includes proteins which play  
20 a role in, *e.g.*, catalyze an enzymatic reaction, in one or more amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside or trehalose metabolic pathways. Examples of MP proteins include those encoded by the MP genes set forth in Table 1 and by the odd-numbered SEQ ID NOs. The terms "MP gene" or "MP nucleic acid sequence" include nucleic acid sequences encoding an MP protein, which consist of a  
25 coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of MP genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (*e.g.*, kg product per hour per liter). The term "efficiency of  
30 production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes

the efficiency of the conversion of the carbon source into the product (*i.e.*, fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (*e.g.*, the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

In another embodiment, the MP molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. Using recombinant genetic techniques, one or more of the biosynthetic or degradative enzymes of the invention for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose may be manipulated such that its function is modulated. For example, a biosynthetic enzyme may be improved in efficiency, or its allosteric control region destroyed such that feedback inhibition of production of the compound is prevented. Similarly, a degradative enzyme may be deleted or modified by substitution, deletion, or addition such that its degradative activity is lessened for the desired compound without impairing the viability of the cell. In each case, the overall yield or rate of production of one of these desired fine chemicals may be increased.

It is also possible that such alterations in the protein and nucleotide molecules of the invention may improve the production of other fine chemicals besides the amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Metabolism of any one compound is necessarily intertwined with other biosynthetic and



degradative pathways within the cell, and necessary cofactors, intermediates, or substrates in one pathway are likely supplied or limited by another such pathway.

Therefore, by modulating the activity of one or more of the proteins of the invention, the production or efficiency of activity of another fine chemical biosynthetic or degradative pathway may be impacted. For example, amino acids serve as the structural units of all proteins, yet may be present intracellularly in levels which are limiting for protein synthesis; therefore, by increasing the efficiency of production or the yields of one or more amino acids within the cell, proteins, such as biosynthetic or degradative proteins, may be more readily synthesized. Likewise, an alteration in a metabolic pathway enzyme such that a particular side reaction becomes more or less favored may result in the over- or under-production of one or more compounds which are utilized as intermediates or substrates for the production of a desired fine chemical.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* MP DNAs and the predicted amino acid sequences of the *C. glutamicum* MP proteins are shown in the Sequence Listing as odd-numbered SEQ ID NOs and even-numbered SEQ ID NOs, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode metabolic pathway proteins.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of the invention (e.g., the sequence of an even-numbered SEQ ID NO of the Sequence Listing). As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The MP protein or a biologically active portion or fragment thereof of the invention can catalyze an enzymatic reaction in one or more amino acid, vitamin,

cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections:

5

*A. Isolated Nucleic Acid Molecules*

One aspect of the invention pertains to isolated nucleic acid molecules that encode MP polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or  
10 amplification of MP-encoding nucleic acid (*e.g.*, MP DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides  
15 of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic  
20 acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank  
25 the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (*e.g.*, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

30

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having a nucleotide sequence of an odd-numbered SEQ ID NO of the Sequence Listing, or a portion thereof, can be isolated using standard molecular biology techniques and the

sequence information provided herein. For example, a *C. glutamicum* MP DNA can be isolated from a *C. glutamicum* library using all or portion of one of the odd-numbered SEQ ID NO sequences of the Sequence Listing as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the nucleic acid sequences of the invention (e.g., an odd-numbered SEQ ID NO:) can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the nucleic acid sequences of the invention (e.g., an odd-numbered SEQ ID NO of the Sequence Listing) can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence). For example, mRNA can be isolated from normal endothelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin *et al.* (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in the Sequence Listing. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an MP nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in the Sequence Listing. The nucleic acid sequences of the invention, as set forth in the Sequence Listing, correspond to the *Corynebacterium glutamicum* MP DNAs of the invention. This DNA comprises sequences encoding MP proteins (i.e., the "coding region", indicated in each odd-numbered SEQ ID NO: sequence in the Sequence Listing), as well as 5' untranslated

sequences and 3' untranslated sequences, also indicated in each odd-numbered SEQ ID NO: in the Sequence Listing. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the nucleic acid sequences of the Sequence Listing.

For the purposes of this application, it will be understood that each of the nucleic acid and amino acid sequences set forth in the Sequence Listing has an identifying RXA, RXN, RXS, or RXC number having the designation "RXA", "RXN", "RXS", or "RXC" followed by 5 digits (*i.e.*, RXA00007, RXN00023, RXS00116, or RXC00128). Each of the nucleic acid sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation "one of the odd-numbered sequences of the Sequence Listing", then, refers to any of the nucleic acid sequences in the Sequence Listing, which may also be distinguished by their differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is also set forth in the Sequence Listing, as an even-numbered SEQ ID NO: immediately following the corresponding nucleic acid sequence. For example, the coding region for RXA02229 is set forth in SEQ ID NO:1, while the amino acid sequence which it encodes is set forth as SEQ ID NO:2. The sequences of the nucleic acid molecules of the invention are identified by the same RXA, RXN, RXS, or RXC designations as the amino acid molecules which they encode, such that they can be readily correlated. For example, the amino acid sequences designated RXA02229, RX00351, RXS02970, and RXC02390 are translations of the coding regions of the nucleotide sequences of nucleic acid molecules RXA02229, RX00351, RXS02970, and RXC02390, respectively. The correspondence between the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention and their assigned SEQ ID NOs is set forth in Table 1.

Several of the genes of the invention are "F-designated genes". An F-designated gene includes those genes set forth in Table 1 which have an 'F' in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:5, designated, as indicated on Table 1, as "F RXA01009", is an F-designated gene, as are SEQ ID NOs: 73, 75, and 77 (designated on Table 1 as "F RXA00007", "F RXA00364", and "F RXA00367", respectively).

In one embodiment, the nucleic acid molecules of the present invention are not intended to include *C. glutamicum* those compiled in Table 2. In the case of the *dapD* gene, a sequence for this gene was published in Wehrmann, A., *et al.* (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing), or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences of the invention is one which is sufficiently complementary to one of the nucleotide sequences shown in the Sequence Listing (*e.g.*, the sequence of an odd-numbered SEQ ID NO:) such that it can hybridize to one of the nucleotide sequences of the invention, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing), or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to one of the nucleotide sequences of the invention, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of the sequence of one of the odd-numbered SEQ ID NOs

of the Sequence Listing, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an MP protein. The nucleotide sequences determined from the cloning of the MP genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning MP  
5 homologues in other cell types and organisms, as well as MP homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one  
10 of the nucleotide sequences of the invention (*e.g.*, a sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing), an anti-sense sequence of one of these sequences, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of the invention can be used in PCR reactions to clone MP homologues. Probes based on the MP nucleotide sequences can be used to detect transcripts or  
15 genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an MP protein, such as by measuring a level of an MP-encoding  
20 nucleic acid in a sample of cells from a subject *e.g.*, detecting MP mRNA levels or determining whether a genomic MP gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-  
25 numbered SEQ ID NO of the Sequence Listing) such that the protein or portion thereof maintains the ability to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or  
30 equivalent (*e.g.*, an amino acid residue which has a similar side chain as an amino acid residue in a sequence of one of the even-numbered SEQ ID NOs of the Sequence Listing) amino acid residues to an amino acid sequence of the invention such that the

protein or portion thereof is able to catalyze an enzymatic reaction in a *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside or trehalose metabolic pathway. Protein members of such metabolic pathways, as described herein, function to catalyze the biosynthesis or degradation of one or more of: amino acids, 5 vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose. Examples of such activities are also described herein. Thus, "the function of an MP protein" contributes to the overall functioning of one or more such metabolic pathway and contributes, either directly or indirectly, to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of MP protein activities are set 10 forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of 15 the Sequence Listing).

Portions of proteins encoded by the MP nucleic acid molecules of the invention are preferably biologically active portions of one of the MP proteins. As used herein, the term "biologically active portion of an MP protein" is intended to include a portion, *e.g.*, a domain/motif, of an MP protein that catalyzes an enzymatic reaction in one or 20 more *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, or has an activity as set forth in Table 1. To determine whether an MP protein or a biologically active portion thereof can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, an assay of enzymatic activity may be performed. Such 25 assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an MP protein can be prepared by isolating a portion of one of the amino acid sequences of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence 30 Listing), expressing the encoded portion of the MP protein or peptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the MP protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing) (and portions thereof) due to degeneracy of the genetic code and thus encode the same MP protein as that encoded by the nucleotide sequences of the invention. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in the Sequence Listing (*e.g.*, an even-numbered SEQ ID NO:). In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of the invention (encoded by an open reading frame shown in an odd-numbered SEQ ID NO: of the Sequence Listing).

It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Tables 2 or 4 which were available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the invention which is greater than that of a sequence of the prior art (*e.g.*, a Genbank sequence (or the protein encoded by such a sequence) set forth in Tables 2 or 4). For example, the invention includes a nucleotide sequence which is greater than and/or at least 40% identical to the nucleotide sequence designated RXA00115 (SEQ ID NO:185), a nucleotide sequence which is greater than and/or at least % identical to the nucleotide sequence designated RXA00131 (SEQ ID NO:991), and a nucleotide sequence which is greater than and/or at least 39% identical to the nucleotide sequence designated RXA00219 (SEQ ID NO:345). One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (*e.g.*, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%,



74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more identical) are also encompassed by the invention.

- In addition to the *C. glutamicum* MP nucleotide sequences set forth in the
- 5 Sequence Listing as odd-numbered SEQ ID NOs, it will be appreciated by one of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of MP proteins may exist within a population (*e.g.*, the *C. glutamicum* population). Such genetic polymorphism in the MP gene may exist among individuals within a population due to natural variation. As used herein, the terms
- 10 "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an MP protein, preferably a *C. glutamicum* MP protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the MP gene. Any and all such nucleotide variations and resulting amino acid
- 15 polymorphisms in MP that are the result of natural variation and that do not alter the functional activity of MP proteins are intended to be within the scope of the invention.

- Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* MP DNA of the invention can be isolated based on their homology to the *C. glutamicum* MP nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard
- 20 hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of an odd-numbered SEQ ID NO: of the Sequence Listing. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or
- 25 more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about
- 30 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to one of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a nucleotide sequence of the invention corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* MP protein.

10 In addition to naturally-occurring variants of the MP sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of the invention, thereby leading to changes in the amino acid sequence of the encoded MP protein, without altering the functional ability of the MP protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a nucleotide sequence of the invention. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the MP proteins (*e.g.*, an even-numbered SEQ ID NO: of the Sequence Listing) without altering the activity of said MP protein, whereas an "essential" amino acid residue is required for MP protein activity.

15 20 Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved in the domain having MP activity) may not be essential for activity and thus are likely to be amenable to alteration without altering MP activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding MP proteins that contain changes in amino acid residues that are not essential for MP activity. Such MP proteins differ in amino acid sequence from a sequence of an even-numbered SEQ ID NO: of the Sequence Listing yet retain at least one of the MP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of the invention and is capable of catalyzing an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic

25 30

acid molecule is at least about 50-60% homologous to the amino acid sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing, more preferably at least about 60-70% homologous to one of these sequences, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of these sequences, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the amino acid sequences of the invention.

To determine the percent homology of two amino acid sequences (*e.g.*, one of the amino acid sequences of the invention and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (*e.g.*, one of the amino acid sequences of the invention) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (*e.g.*, a mutant form of the amino acid sequence), then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an MP protein homologous to a protein sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of the invention such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the nucleotide sequences of the invention by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic

acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an MP protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an MP activity described herein to identify mutants that retain MP activity. Following mutagenesis of the nucleotide sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding MP proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded DNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire MP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MP protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the entire coding region of SEQ ID NO. 1 (RXA02229) comprises nucleotides 1 to 825). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MP. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding MP disclosed herein (*e.g.*, the sequences set forth as odd-numbered SEQ ID NOs in the Sequence Listing), antisense

nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MCT protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by  
5 conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecule to a peptide or  
10 an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

15 In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a  
20 chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they  
25 have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave MP mRNA transcripts to thereby inhibit translation of MP mRNA. A ribozyme having specificity for an MP-encoding nucleic acid can be designed based upon the nucleotide sequence of an MP DNA disclosed herein (*i.e.*, SEQ ID NO: 1  
30 (RXA02229). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MP-encoding mRNA. See, *e.g.*, Cech *et al.*

U.S. Patent No. 4,987,071 and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, MP mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

- 5           Alternatively, MP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an MP nucleotide sequence (*e.g.*, an MP promoter and/or enhancers) to form triple helical structures that prevent transcription of an MP gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and
- 10   Maher, L.J. (1992) *Bioassays* 14(12):807-15.

#### *B. Recombinant Expression Vectors and Host Cells*

- Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MP protein (or a portion thereof). As
- 15   used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of
- 20   autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to
- 25   which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector.
- However, the invention is intended to include such other forms of expression vectors,
- 30   such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

- The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is
- 5 operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory
- 10 sequence" is intended to include promoters, repressor binding sites, activator binding sites, enhancers and other expression control elements (*e.g.*, terminators, polyadenylation signals, or other elements of mRNA secondary structure). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990).
- 15 Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. Preferred regulatory sequences are, for example, promoters such as *cos*-, *tac*-, *trp*-, *tet*-, *trp-tet*-, *lpp*-, *lac*-, *lpp-lac*-, *lacI<sup>q</sup>*-, *T7*-, *T5*-, *T3*-, *gal*-, *trc*-, *ara*-, *SP6*-, *amy*, *SPO2*,  $\lambda$ -*P<sub>R</sub>*- or  $\lambda$  *P<sub>L</sub>*, which are used preferably in bacteria.
- 20 Additional regulatory sequences are, for example, promoters from yeasts and fungi, such as *ADC1*, *MF $\alpha$* , *AC*, *P-60*, *CYC1*, *GAPDH*, *TEF*, *rp28*, *ADH*, promoters from plants such as *CaMV/35S*, *SSU*, *OCS*, *lib4*, *usp*, *STLS1*, *B33*, *nos* or *ubiquitin*- or *phaseolin*-promoters. It is also possible to use artificial promoters. It will be appreciated by one of ordinary skill in the art that the design of the expression vector can depend on such
- 25 factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, MP proteins, mutant forms of MP proteins, fusion proteins, etc.).
- 30 The recombinant expression vectors of the invention can be designed for expression of MP proteins in prokaryotic or eukaryotic cells. For example, MP genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus



expression vectors), yeast and other fungal cells (see Romanos, M.A. *et al.* (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. *et al.* (1991) "Heterologous gene expression in filamentous fungi" in: *More Gene Manipulations in Fungi*, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: *Applied Molecular Genetics of Fungi*, Peberdy, J.F. *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* -mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the MP protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from

the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant MP protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

- 5           Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III113-B1,  $\lambda$ gt11, pBdCl, and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89; and
- 10   Pouwels *et al.*, eds. (1985) *Cloning Vectors*. Elsevier: New York ISBN 0 444 904018). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by
- 15   host strains BL21(DE3) or HMS174(DE3) from a resident  $\lambda$  prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter. For transformation of other varieties of bacteria, appropriate vectors may be selected. For example, the plasmids pIJ101, pIJ364, pIJ702 and pIJ361 are known to be useful in transforming *Streptomyces*, while plasmids pUB110, pC194, or pBD214 are suited for transformation
- 20   of *Bacillus* species. Several plasmids of use in the transfer of genetic information into *Corynebacterium* include pHM1519, pBL1, pSA77, or pAJ667 (Pouwels *et al.*, eds. (1985) *Cloning Vectors*. Elsevier: New York ISBN 0 444 904018).

- One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the
- 25   recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum*
- 30   (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the MP protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), , 2  $\mu$ , pAG-1, Yep6, Yep13, pEMBLYe23, pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, *et al.*, eds., p. 1-28, Cambridge University Press: Cambridge, and Pouwels *et al.*, eds. (1985) *Cloning Vectors*. Elsevier: New York (IBSN 0 444 904018).

Alternatively, the MP proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the MP proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (*e.g.*, the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl. Acid. Res.* 12: 8711-8721, and include pLGV23, pGHIac+, pBIN19, pAK2004, and pDH51 (Pouwels *et al.*, eds. (1985) *Cloning Vectors*. Elsevier: New York IBSN 0 444 904018).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both

prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

- 5 In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* 10 (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), 15 pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) 20 *Genes Dev.* 3:537-546).

- The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an 25 RNA molecule which is antisense to MP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA.
- 30 The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell

type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an MP protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those of ordinary skill in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, linear DNA or RNA (*e.g.*, a linearized vector or a gene construct alone without a vector) or nucleic acid in the form of a vector (*e.g.*, a plasmid, phage, phasmid, phagemid, transposon or other DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these

integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be  
5 introduced into a host cell on the same vector as that encoding an MP protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which  
10 contains at least a portion of an MP gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the MP gene. Preferably, this MP gene is a *Corynebacterium glutamicum* MP gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous  
15 recombination, the endogenous MP gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous MP gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the  
20 endogenous MP protein). In the homologous recombination vector, the altered portion of the MP gene is flanked at its 5' and 3' ends by additional nucleic acid of the MP gene to allow for homologous recombination to occur between the exogenous MP gene carried by the vector and an endogenous MP gene in a microorganism. The additional flanking MP nucleic acid is of sufficient length for successful homologous  
25 recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see *e.g.*, Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (*e.g.*, by electroporation) and cells in which the introduced MP gene has homologously recombined with the  
30 endogenous MP gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene.

For example, inclusion of an MP gene on a vector placing it under control of the lac operon permits expression of the MP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

In another embodiment, an endogenous MP gene in a host cell is disrupted (*e.g.*,  
5 by homologous recombination or other genetic means known in the art) such that expression of its protein product does not occur. In another embodiment, an endogenous or introduced MP gene in a host cell has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional MP protein. In still another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or  
10 inducer) of an MP gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the MP gene is modulated. One of ordinary skill in the art will appreciate that host cells containing more than one of the described MP gene and protein modifications may be readily produced using the methods of the invention, and are meant to be included in the present  
15 invention.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an MP protein. Accordingly, the invention further provides methods for producing MP proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of  
20 invention (into which a recombinant expression vector encoding an MP protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered MP protein) in a suitable medium until MP protein is produced. In another embodiment, the method further comprises isolating MP proteins from the medium or the host cell.

25

### *C. Isolated MP Proteins*

Another aspect of the invention pertains to isolated MP proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA  
30 techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of MP protein in which the protein is separated from cellular components of the cells in which

it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of MP protein having less than about 30% (by dry weight) of non-MP protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-MP protein, still  
5 more preferably less than about 10% of non-MP protein, and most preferably less than about 5% non-MP protein. When the MP protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The  
10 language "substantially free of chemical precursors or other chemicals" includes preparations of MP protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of MP protein having less than about 30% (by dry weight) of chemical  
15 precursors or non-MP chemicals, more preferably less than about 20% chemical precursors or non-MP chemicals, still more preferably less than about 10% chemical precursors or non-MP chemicals, and most preferably less than about 5% chemical precursors or non-MP chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism  
20 from which the MP protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* MP protein in a microorganism such as *C. glutamicum*.

An isolated MP protein or a portion thereof of the invention can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide,  
25 nucleoside, or trehalose metabolic pathway, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) such that the protein or portion thereof maintains the ability to catalyze an enzymatic  
30 reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an MP protein of



the invention has an amino acid sequence set forth as an even-numbered SEQ ID NO: of the Sequence Listing. In yet another preferred embodiment, the MP protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing). In still another preferred embodiment, the MP protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to one of the nucleic acid sequences of the invention, or a portion thereof. Ranges and identity values intermediate to the above-recited values, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. The preferred MP proteins of the present invention also preferably possess at least one of the MP activities described herein. For example, a preferred MP protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to a nucleotide sequence of the invention, and which can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or which has one or more of the activities set forth in Table 1.

In other embodiments, the MP protein is substantially homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) and retains the functional activity of the protein of one of the amino acid sequences of the invention yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the MP protein is a protein which comprises an amino acid sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%,

78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of the invention and which has at least one of the MP activities described herein. Ranges and identity values  
5 intermediate to the above-recited values, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire  
10 amino acid sequence of the invention.

Biologically active portions of an MP protein include peptides comprising amino acid sequences derived from the amino acid sequence of an MP protein, e.g., an amino acid sequence of an even-numbered SEQ ID NO: of the Sequence Listing or the amino acid sequence of a protein homologous to an MP protein, which include fewer amino  
15 acids than a full length MP protein or the full length protein which is homologous to an MP protein, and exhibit at least one activity of an MP protein. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an MP protein. Moreover, other biologically active portions, in  
20 which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an MP protein include one or more selected domains/motifs or portions thereof having biological activity.

MP proteins are preferably produced by recombinant DNA techniques. For  
25 example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the MP protein is expressed in the host cell. The MP protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an MP protein,  
30 polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native MP protein can be isolated from cells (e.g., endothelial

cells), for example using an anti-MP antibody, which can be produced by standard techniques utilizing an MP protein or fragment thereof of this invention.

The invention also provides MP chimeric or fusion proteins. As used herein, an MP "chimeric protein" or "fusion protein" comprises an MP polypeptide operatively  
5 linked to a non-MP polypeptide. An "MP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to MP, whereas a "non-MP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the MP protein, *e.g.*, a protein which is different from the MP protein and which is derived from the same or a different organism. Within the  
10 fusion protein, the term "operatively linked" is intended to indicate that the MP polypeptide and the non-MP polypeptide are fused in-frame to each other. The non-MP polypeptide can be fused to the N-terminus or C-terminus of the MP polypeptide. For example, in one embodiment the fusion protein is a GST-MP fusion protein in which the MP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins  
15 can facilitate the purification of recombinant MP proteins. In another embodiment, the fusion protein is an MP protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of an MP protein can be increased through use of a heterologous signal sequence.

Preferably, an MP chimeric or fusion protein of the invention is produced by  
20 standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid  
25 undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric  
30 gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An MP-

encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MP protein.

Homologues of the MP protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the MP protein. As used herein, the term "homologue" refers to a variant form of the MP protein which acts as an agonist or antagonist of the activity of the MP protein. An agonist of the MP protein can retain substantially the same, or a subset, of the biological activities of the MP protein. An antagonist of the MP protein can inhibit one or more of the activities of the naturally occurring form of the MP protein, by, for example, competitively binding to a downstream or upstream member of the MP cascade which includes the MP protein. Thus, the *C. glutamicum* MP protein and homologues thereof of the present invention may modulate the activity of one or more metabolic pathways in which MP proteins play a role in this microorganism.

In an alternative embodiment, homologues of the MP protein can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the MP protein for MP protein agonist or antagonist activity. In one embodiment, a variegated library of MP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of MP variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential MP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of MP sequences therein. There are a variety of methods which can be used to produce libraries of potential MP homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential MP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the MP protein coding can be used to generate a variegated population of MP fragments for screening and subsequent selection of homologues of an MP protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an MP coding sequence with a nuclease under conditions wherein nicking occurs only  
5 about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression  
10 vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the MP protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA  
libraries for gene products having a selected property. Such techniques are adaptable for  
15 rapid screening of the gene libraries generated by the combinatorial mutagenesis of MP homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of  
vectors, and expressing the combinatorial genes under conditions in which detection of a  
20 desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MP homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

25 In another embodiment, cell based assays can be exploited to analyze a variegated MP library, using methods well known in the art.

#### *D. Uses and Methods of the Invention*

The nucleic acid molecules, proteins, protein homologues, fusion proteins,  
30 primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C.*

*glutamicum* sequences of interest; evolutionary studies; determination of MP protein regions required for function; modulation of an MP protein activity; modulation of the activity of an MP pathway; and modulation of cellular production of a desired compound, such as a fine chemical.

- 5           The MP nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the
- 10   extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is not pathogenic to humans, it is related to species which are human pathogens, such as *Corynebacterium diphtheriae*.
- 15   *Corynebacterium diphtheriae* is the causative agent of diphtheria, a rapidly developing, acute, febrile infection which involves both local and systemic pathology. In this disease, a local lesion develops in the upper respiratory tract and involves necrotic injury to epithelial cells; the bacilli secrete toxin which is disseminated through this lesion to distal susceptible tissues of the body. Degenerative changes brought about by the
- 20   inhibition of protein synthesis in these tissues, which include heart, muscle, peripheral nerves, adrenals, kidneys, liver and spleen, result in the systemic pathology of the disease. Diphtheria continues to have high incidence in many parts of the world, including Africa, Asia, Eastern Europe and the independent states of the former Soviet Union. An ongoing epidemic of diphtheria in the latter two regions has resulted in at
- 25   least 5,000 deaths since 1990.

          In one embodiment, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (e.g., the sequences set forth as odd-numbered or even-numbered SEQ ID NOs, respectively, in

30   the Sequence Listing) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject. *C. glutamicum* and *C. diphtheriae* are related bacteria, and many of the nucleic acid and protein molecules in *C. glutamicum*

are homologous to *C. diphtheriae* nucleic acid and protein molecules, and can therefore be used to detect *C. diphtheriae* in a subject.

The nucleic acid and protein molecules of the invention may also serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to  
5 identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable  
10 labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that  
15 these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The MP nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells;  
20 by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the  
25 enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the MP nucleic acid molecules of the invention may result in the production of MP proteins having functional differences from the wild-type MP  
30 proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

The invention also provides methods for screening molecules which modulate the activity of an MP protein, either by interacting with the protein itself or a substrate or binding partner of the MP protein, or by modulating the transcription or translation of an MP nucleic acid molecule of the invention. In such methods, a microorganism  
5 expressing one or more MP proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of the MP protein is assessed.

When the desired fine chemical to be isolated from large-scale fermentative culture of *C. glutamicum* is an amino acid, a vitamin, a cofactor, a nutraceutical, a  
10 nucleotide, a nucleoside, or trehalose, modulation of the activity or efficiency of activity of one or more of the proteins of the invention by recombinant genetic mechanisms may directly impact the production of one of these fine chemicals. For example, in the case of an enzyme in a biosynthetic pathway for a desired amino acid, improvement in efficiency or activity of the enzyme (including the presence of multiple copies of the  
15 gene) should lead to an increased production or efficiency of production of that desired amino acid. In the case of an enzyme in a biosynthetic pathway for an amino acid whose synthesis is in competition with the synthesis of a desired amino acid, any decrease in the efficiency or activity of this enzyme (including deletion of the gene) should result in an increase in production or efficiency of production of the desired amino acid, due to  
20 decreased competition for intermediate compounds and/or energy. In the case of an enzyme in a degradation pathway for a desired amino acid, any decrease in efficiency or activity of the enzyme should result in a greater yield or efficiency of production of the desired product due to a decrease in its degradation. Lastly, mutagenesis of an enzyme involved in the biosynthesis of a desired amino acid such that this enzyme is no longer is  
25 capable of feedback inhibition should result in increased yields or efficiency of production of the desired amino acid. The same should apply to the biosynthetic and degradative enzymes of the invention involved in the metabolism of vitamins, cofactors, nutraceuticals, nucleotides, nucleosides and trehalose.

Similarly, when the desired fine chemical is not one of the aforementioned  
30 compounds, the modulation of activity of one of the proteins of the invention may still impact the yield and/or efficiency of production of the compound from large-scale culture of *C. glutamicum*. The metabolic pathways of any organism are closely



interconnected; the intermediate used by one pathway is often supplied by a different pathway. Enzyme expression and function may be regulated based on the cellular levels of a compound from a different metabolic process, and the cellular levels of molecules necessary for basic growth, such as amino acids and nucleotides, may critically affect the viability of the microorganism in large-scale culture. Thus, modulation of an amino acid biosynthesis enzyme, for example, such that it is no longer responsive to feedback inhibition or such that it is improved in efficiency or turnover may result in increased cellular levels of one or more amino acids. In turn, this increased pool of amino acids provides not only an increased supply of molecules necessary for protein synthesis, but also of molecules which are utilized as intermediates and precursors in a number of other biosynthetic pathways. If a particular amino acid had been limiting in the cell, its increased production might increase the ability of the cell to perform numerous other metabolic reactions, as well as enabling the cell to more efficiently produce proteins of all kinds, possibly increasing the overall growth rate or survival ability of the cell in large scale culture. Increased viability improves the number of cells capable of producing the desired fine chemical in fermentative culture, thereby increasing the yield of this compound. Similar processes are possible by the modulation of activity of a degradative enzyme of the invention such that the enzyme no longer catalyzes, or catalyzes less efficiently, the degradation of a cellular compound which is important for the biosynthesis of a desired compound, or which will enable the cell to grow and reproduce more efficiently in large-scale culture. It should be emphasized that optimizing the degradative activity or decreasing the biosynthetic activity of certain molecules of the invention may also have a beneficial effect on the production of certain fine chemicals from *C. glutamicum*. For example, by decreasing the efficiency of activity of a biosynthetic enzyme in a pathway which competes with the biosynthetic pathway of a desired compound for one or more intermediates, more of those intermediates should be available for conversion to the desired product. A similar situation may call for the improvement of degradative ability or efficiency of one or more proteins of the invention.

30 This aforementioned list of mutagenesis strategies for MP proteins to result in increased yields of a desired compound is not meant to be limiting; variations on these mutagenesis strategies will be readily apparent to one of ordinary skill in the art. By

these mechanisms, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated MP nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any

5 natural product of *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not

10 be construed as limiting. The contents of all references, patent applications, patents, published patent applications, Tables, and the sequence listing cited throughout this application are hereby incorporated by reference.

Table 1 (continued)

## Alanine and Aspartate and Asparagine metabolism

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
109	110	RXA02139	GR00639	6739	4901	ASPARAGINE SYNTHETASE (GLUTAMINE-HYDROLYZING) (EC 6.3.5.4)
111	112	RXN00116	VV0100	26974	25814	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
113	114	F RXA00116	GR00018	510	4	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
115	116	RXN00618	VV0135	10288	9182	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
117	118	F RXA00618	GR00163	213	746	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
119	120	F RXA00627	GR00164	854	1138	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
121	122	RXA02550	GR00729	1585	275	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
123	124	RXA02193	GR00645	1942	365	ASPARTATE AMONIA-LYASE (EC 4.3.1.1)
125	126	RXA02432	GR00708	2669	1695	L-ASPARAGINASE (EC 3.5.1.1)
127	128	RXN03003	VV0138	680	6	ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1)
129	130	RXN00508	VV0086	4701	5783	ALANINE RACEMASE (EC 5.1.1.1)
131	132	RXN00636	VV0135	20972	19944	ALANINE RACEMASE, BIOSYNTHETIC (EC 5.1.1.1)

## beta-Alanine metabolism

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
133	134	RXA02536	GR00726	8581	7826	BETA-UREIDOPROPIONASE (EC 3.5.1.6)
135	136	RXS00870				METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE (ACYLATING) (EC 1.2.1.27)
137	138	RXS02299				ASPARTATE 1-DECARBOXYLASE PRECURSOR (EC 4.1.1.11)

## Glycine and serine metabolism

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
139	140	RXA01561	GR00435	1113	2042	L-SERINE DEHYDRATASE (EC 4.2.1.13)
141	142	RXA01850	GR00525	481	1827	L-SERINE DEHYDRATASE (EC 4.2.1.13)
143	144	RXA00580	GR00156	7343	6042	SERINE HYDROXYMETHYLTRANSFERASE (EC 2.1.2.1)
145	146	RXA01821	GR00515	10253	9876	SARCOSINE OXIDASE (EC 1.5.3.1)
147	148	RXN02263	VV0202	11783	12160	SARCOSINE OXIDASE (EC 1.5.3.1)
149	150	F RXA02263	GR00654	33454	33813	SARCOSINE OXIDASE (EC 1.5.3.1)
151	152	RXA02176	GR00641	11454	12581	PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52)
153	154	RXN02758	GR00766	5082	4648	PHOSPHOSERINE PHOSPHATASE (EC 3.1.3.3)
155	156	F RXA02479	GR00717	393	4	PHOSPHOSERINE PHOSPHATASE (EC 3.1.3.3)
157	158	F RXA02758	GR00766	5082	4648	PHOSPHOSERINE PHOSPHATASE (EC 3.1.3.3)
159	160	F RXA02759	GR00766	5330	5220	PHOSPHOSERINE PHOSPHATASE (EC 3.1.3.3)
161	162	RXA02501	GR00720	15041	13977	PHOSPHOSERINE PHOSPHATASE (EC 3.1.3.3)
163	164	RXN03105	VV0074	15857	15423	SARCOSINE OXIDASE (EC 1.5.3.1)
165	166	RXS01130				D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95)
167	168	RXS03112				D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95)

Table 1 (continued)

## Threonine metabolism

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
169	170	RXN00969	VV0149	12053	13387	HOMOSERINE DEHYDROGENASE (EC 1.1.1.3)
171	172	F RXA00974	GR00274	2623	3015	HOMOSERINE DEHYDROGENASE (EC 1.1.1.3)
173	174	RXA00970	GR00273	161	1087	HOMOSERINE KINASE (EC 2.7.1.39)
175	176	RXA00330	GR00057	12968	14410	THREONINE SYNTHASE (EC 4.2.99.2)
177	178	RXN00403	VV0086	70041	68911	HOMOSERINE O-ACETYLTRANSFERASE
179	180	F RXA00403	GR00088	723	1832	HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.11)
181	182	RXC01207				CYTOSOLIC PROTEIN INVOLVED IN METABOLISM OF LYSINE AND THREONINE
183	184	RXC00152				MEMBRANE ASSOCIATED PROTEIN INVOLVED IN THREONINE METABOLISM

## Metabolism of methionine and S-adenosyl methionine

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
185	186	RXA00115	GR00017	5359	4313	HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31)
187	188	RXN00403	VV0086	70041	68911	HOMOSERINE O-ACETYLTRANSFERASE
189	190	F RXA00403	GR00088	723	1832	HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.11)
191	192	RXS03158	GR00038	2404	1811	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
193	194	F RXA00254	GR00252	3085	2039	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
195	196	RXA02532	GR00726	1919	2521	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
197	198	RXS03159	GR00770	16286	15297	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
199	200	F RXA02768	GR00032	70787	70188	CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9)
201	202	RXA00216	VV0086			5-methyltetrahydrofolate-homocysteine methyltransferase (methionine synthetase)
203	204	RXN00402	GR00088	1	576	O-ACETYLHOMOSERINE SULFHYDRYLASE (EC 4.2.99.10) / O-ACETYL SERINE SULFHYDRYLASE (EC 4.2.99.8)
205	206	F RXA00402	GR00088	1	576	O-ACETYLHOMOSERINE SULFHYDRYLASE (EC 4.2.99.10) / O-ACETYL SERINE SULFHYDRYLASE (EC 4.2.99.8)
207	208	RXA00405	GR00089	3289	3801	O-ACETYLHOMOSERINE SULFHYDRYLASE (EC 4.2.99.10) / O-ACETYL SERINE SULFHYDRYLASE (EC 4.2.99.8)
209	210	RXA02197	GR00845	4552	4025	5-METHYLTETRAHYDROFOLATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.13)
211	212	RXN02198	VV0302	9228	11726	5-METHYLTETRAHYDROFOLATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.13)
213	214	F RXA02198	GR00846	2483	6	5-METHYLTETRAHYDROFOLATE--HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.13)
215	216	RXN03074	VV0042	2238	1741	S-ADENOSYLMETHIONINE:2-DEMETHYLMENAQUINONE METHYLTRANSFERASE (EC 2.1.-.-)
217	218	F RXA02906	GR10044	1142	645	S-ADENOSYLMETHIONINE:2-DEMETHYLMENAQUINONE METHYLTRANSFERASE (EC 2.1.-.-)
219	220	RXN00132	VV0124	3612	5045	ADENOSYLMETHIONINE:2-DEMETHYLMENAQUINONE METHYLTRANSFERASE (EC 3.3.1.1)
221	222	F RXA00132	GR00020	7728	7624	ADENOSYLMETHIONINE:2-DEMETHYLMENAQUINONE METHYLTRANSFERASE (EC 3.3.1.1)

- 133 -

min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

5

**Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.**

Using DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (see e.g., Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) *Proc. Natl. Acad. Sci. USA*, 75:3737-3741); pACYC177 (Change & Cohen (1978) *J. Bacteriol* 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) *Gene* 53:283-286. Gene libraries specifically for use in *C. glutamicum* may be constructed using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

20

**Example 3: DNA Sequencing and Computational Functional Analysis**

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R.D. *et al.* (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., *Science*, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

25

**Example 4: *In vivo* Mutagenesis**

*In vivo* mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain

30

the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those of ordinary skill in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

**Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum***

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J.F. *et al.* (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin of replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli* and *C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. *et al.* (1985) *J. Bacteriol.* 162:591-597, Martin J.F. *et al.* (1987) *Biotechnology*, 5:137-146 and Eikmanns, B.J. *et al.* (1991) *Gene*, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. *et al.* (1984) *J. Bacteriol.* 159:306-311), electroporation (Liebl, E. *et al.* (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A *et al.*

is also possible to maintain a constant culture pH through the addition of NaOH or NH<sub>4</sub>OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the micro-organisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD<sub>600</sub> of 0.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2.5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

25

#### **Example 8 – *In vitro* Analysis of the Function of Mutant Proteins**

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one of ordinary skill in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be

30

advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or  
5 ammonia salts, such as  $\text{NH}_4\text{Cl}$  or  $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{NH}_4\text{OH}$ , nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt,  
10 molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic  
15 acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A  
20 Practical Approach (*eds.* P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFCO) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if  
25 necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to  
30 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It



found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) *Enzymes*. Longmans: London; Fersht, (1985) *Enzyme Structure and Mechanism*. Freeman: New York; Walsh, (1979) *Enzymatic Reaction Mechanisms*. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) *Fundamentals of Enzymology*. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) *The Enzymes*, 3<sup>rd</sup> ed. Academic Press: New York; Bisswanger, H., (1994) *Enzymkinetik*, 2<sup>nd</sup> ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) *Methods of Enzymatic Analysis*, 3<sup>rd</sup> ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. *et al.* (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in *Biomembranes, Molecular Structure and Function*, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

#### **Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product**

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (*i.e.*, an amino acid). Such analysis techniques are well known to one of ordinary skill in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example,

Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. *et al.*, (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm *et al.* (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page  
5 469-714, VCH: Weinheim; Belter, P.A. *et al.* (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow,  
10 F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to  
15 determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (*e.g.*, sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these  
20 measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

#### **Example 10: Purification of the Desired Product from *C. glutamicum* Culture**

25 Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and  
30 the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum*

What is claimed:

1. An isolated nucleic acid molecule from *Corynebacterium glutamicum* encoding a metabolic pathway protein, or a portion thereof, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
2. The isolated nucleic acid molecule of claim 1, wherein said metabolic pathway protein is selected from the group consisting of proteins involved in the metabolism of an amino acid, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose.
3. An isolated *Corynebacterium glutamicum* nucleic acid molecule selected from the group consisting of those sequences set forth as odd-numbered SEQ ID NOs of the Sequence Listing, or a portion thereof, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
4. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth as even-numbered SEQ ID NOs of the Sequence Listing, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
5. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth as even-numbered SEQ ID NOs of the Sequence Listing, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
6. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth as odd-numbered SEQ ID NOs of the Sequence Listing, or

a portion thereof, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.

7. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth as odd-numbered SEQ ID NOs of the Sequence Listing, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
8. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1-7 under stringent conditions.
9. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-8 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.
10. A vector comprising the nucleic acid molecule of any one of claims 1-9.
11. The vector of claim 10, which is an expression vector.
12. A host cell transfected with the expression vector of claim 11.
13. The host cell of claim 12, wherein said cell is a microorganism.
14. The host cell of claim 13, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.
15. The host cell of claim 12, wherein the expression of said nucleic acid molecule results in the modulation in production of a fine chemical from said cell.
16. The host cell of claim 15, wherein said fine chemical is selected from the group consisting of: organic acids, nonproteinogenic amino acids, purine and pyrimidine

bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides, and enzymes.

17. A method of producing a polypeptide comprising culturing the host cell of claim 12  
5 in an appropriate culture medium to, thereby, produce the polypeptide.
18. An isolated metabolic pathway polypeptide from *Corynebacterium glutamicum*, or a portion thereof.
- 10 19. The protein of claim 18, wherein said polypeptide is selected from the group of metabolic pathway proteins which participate in the metabolism of an amino acid, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose.
- 15 20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth as even-numbered SEQ ID NOs of the Sequence Listing, provided that the amino acid sequence is not encoded by any of the F-designated genes set forth in Table 1.
- 20 21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth as even-numbered SEQ ID NOs of the Sequence Listing, or a portion thereof, provided that the amino acid sequence is not encoded by any of the F-designated genes set forth in Table 1.
- 25 22. The isolated polypeptide of any of claims 18-21, further comprising heterologous amino acid sequences.
- 30 23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth as odd-numbered SEQ ID NOs of the Sequence Listing, provided that the nucleic acid molecule does not consist of any of the F-designated nucleic acid molecules set forth in Table 1.

24. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth as even-numbered SEQ ID NOs of the Sequence Listing,  
5 provided that the amino acid sequence is not encoded by any of the F-designated genes set forth in Table 1.
25. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 12 such that the fine chemical is produced.  
10
26. The method of claim 25, wherein said method further comprises the step of recovering the fine chemical from said culture.
27. The method of claim 25, wherein said method further comprises the step of  
15 transfecting said cell with the vector of claim 11 to result in a cell containing said vector.
28. The method of claim 25, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.  
20
29. The method of claim 25, wherein said cell is selected from the group consisting of: *Corynebacterium glutamicum*, *Corynebacterium herculis*, *Corynebacterium lilium*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*, *Corynebacterium acetophilum*, *Corynebacterium ammoniagenes*, *Corynebacterium fujiokense*, *Corynebacterium nitrilophilus*, *Brevibacterium ammoniagenes*,  
25 *Brevibacterium butanicum*, *Brevibacterium divaricatum*, *Brevibacterium flavum*, *Brevibacterium healii*, *Brevibacterium ketoglutamicum*, *Brevibacterium ketosoreductum*, *Brevibacterium lactofermentum*, *Brevibacterium linens*, *Brevibacterium paraffinolyticum*, and those strains set forth in Table 3.
30. The method of claim 25, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.  
30

31. The method of claim 25, wherein said fine chemical is selected from the group consisting of: organic acids, nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides, and enzymes.
32. The method of claim 25, wherein said fine chemical is an amino acid.
33. The method of claim 32, wherein said amino acid is drawn from the group consisting of: lysine, glutamate, glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.
34. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-9.
35. A method for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject, comprising detecting the presence of one or more of SEQ ID NOs 1 through 1156 of the Sequence Listing in the subject, provided that the sequences are not or are not encoded by any of the F-designated sequences set forth in Table 1, thereby diagnosing the presence or activity of *Corynebacterium diphtheriae* in the subject.
36. A host cell comprising a nucleic acid molecule selected from the group consisting of the nucleic acid molecules set forth as odd-numbered SEQ ID NOs of the Sequence Listing, wherein the nucleic acid molecule is disrupted.
37. A host cell comprising a nucleic acid molecule selected from the group consisting of the nucleic acid molecules set forth as odd-numbered SEQ ID NOs in the Sequence Listing, wherein the nucleic acid molecule comprises one or more nucleic acid

modifications from the sequence set forth as odd-numbered SEQ ID NOs of the Sequence Listings.

38. A host cell comprising a nucleic acid molecule selected from the group consisting of  
5 the nucleic acid molecules set forth as odd-numbered SEQ ID NOs of the Sequence Listing, wherein the regulatory region of the nucleic acid molecule is modified relative to the wild-type regulatory region of the molecule.



35 40 45  
 Ile Thr Asn Gly Ala Arg Leu Glu Thr Tyr Val Ile Val Gly Asp Ala  
 50 55 60  
 Gly Thr Gly Asn Ile Cys Ile Asn Gly Ala Ala Ala His Leu Ile Asn  
 65 70 75 80  
 Pro Gly Asp Leu Val Ile Ile Met Ser Tyr Leu Gln Ala Thr Asp Ala  
 85 90 95  
 Glu Ala Lys Ala Tyr Glu Pro Lys Ile Val His Val Asp Ala Asp Asn  
 100 105 110  
 Arg Ile Val Ala Leu Gly Asn Asp Leu Ala Glu Ala Leu Pro Gly Ser  
 115 120 125  
 Gly Leu Leu Thr Ser Arg Ser Ile  
 130 135  
  
 <210> 139  
 <211> 1053  
 <212> DNA  
 <213> *Corynebacterium glutamicum*  
  
 <220>  
 <221> CDS  
 <222> (101)..(1030)  
 <223> RXA01561  
  
 <400> 139  
 gtgccagaa attctgcttg cactcaccca agccgttttag caaattgaac ctcacgttca 60  
 taataatggt cattttcatc gagttctaga aaacacaggc atg ctc acc ctc aac 115  
 Met Leu Thr Leu Asn  
 1 5  
 gat gtc atc acc gcc caa caa cga acc gcc cct cat gtt cga cga acg 163  
 Asp Val Ile Thr Ala Gln Gln Arg Thr Ala Pro His Val Arg Arg Thr  
 10 15 20  
 cca ctt ttc gaa gca gac ccc atc gac ggc aca caa atc tgg atc aaa 211  
 Pro Leu Phe Glu Ala Asp Pro Ile Asp Gly Thr Gln Ile Trp Ile Lys  
 25 30 35  
 gca gag ttc ctc caa aag tgc ggc gtg ttc aaa acg cgt gga gca ttc 259  
 Ala Glu Phe Leu Gln Lys Cys Gly Val Phe Lys Thr Arg Gly Ala Phe  
 40 45 50  
 aac cgc cag ctc gca gct tcg gaa aac gga cta ctc gac cca acg gtt 307  
 Asn Arg Gln Leu Ala Ala Ser Glu Asn Gly Leu Leu Asp Pro Thr Val  
 55 60 65  
 ggc atc gtc gcg gca tca ggc gga aac gca gga ctc gca aat gct ttt 355  
 Gly Ile Val Ala Ala Ser Gly Gly Asn Ala Gly Leu Ala Asn Ala Phe  
 70 75 80 85  
 gcc gca gca tcc tta agc gtt ccc gcc acg gta ttg gtg ccc gaa act 403  
 Ala Ala Ala Ser Leu Ser Val Pro Ala Thr Val Leu Val Pro Glu Thr  
 90 95 100

This Page Blank (uspic)

gcc cca caa gta aaa gtt gat cgc ctc aag caa tac ggt gca acc gtg 451  
 Ala Pro Gln Val Lys Val Asp Arg Leu Lys Gln Tyr Gly Ala Thr Val  
 105 110 115

caa caa atc gga tct gaa tat gcg gaa gca ttt gag gca gct caa acc 499  
 Gln Gln Ile Gly Ser Glu Tyr Ala Glu Ala Phe Glu Ala Ala Gln Thr  
 120 125 130

ttt gag tcg gaa act ggt gct ctg ttt tgc cac gcc tac gac cag ccc 547  
 Phe Glu Ser Glu Thr Gly Ala Leu Phe Cys His Ala Tyr Asp Gln Pro  
 135 140 145

gac atc gca gct gga gca ggc gtc att ggg cta gaa att gtc gaa gat 595  
 Asp Ile Ala Ala Gly Ala Gly Val Ile Gly Leu Glu Ile Val Glu Asp  
 150 155 160 165

ctt ccc gac gtt gac acc atc gtg gtt gct gtc ggt ggc ggt gga ctc 643  
 Leu Pro Asp Val Asp Thr Ile Val Val Ala Val Gly Gly Gly Gly Leu  
 170 175 180

tat gca gga atc gca gcc gtc gta gca gcc cac gac atc aaa gtg gtg 691  
 Tyr Ala Gly Ile Ala Ala Val Val Ala Ala His Asp Ile Lys Val Val  
 185 190 195

gcc gtt gaa ccc tcc aaa att cca acc ctg cac aac tca ctc att gcc 739  
 Ala Val Glu Pro Ser Lys Ile Pro Thr Leu His Asn Ser Leu Ile Ala  
 200 205 210

ggc caa cca gtc gat gtg aac gtt tct ggt atc gcg gca gat tct ttg 787  
 Gly Gln Pro Val Asp Val Asn Val Ser Gly Ile Ala Ala Asp Ser Leu  
 215 220 225

ggg gct cgc caa att gga cga gaa gcc ttt gac atc gca act gcc cat 835  
 Gly Ala Arg Gln Ile Gly Arg Glu Ala Phe Asp Ile Ala Thr Ala His  
 230 235 240 245

ccc cca ata ggc gtc cta gtg gac gat gaa gca atc atc gca gct cga 883  
 Pro Pro Ile Gly Val Leu Val Asp Asp Glu Ala Ile Ile Ala Ala Arg  
 250 255 260

cgc cac ctc tgg gac aac tac cgc atc cct gcc gag cat ggc gct gcc 931  
 Arg His Leu Trp Asp Asn Tyr Arg Ile Pro Ala Glu His Gly Ala Ala  
 265 270 275

gca gca ctc gcc tct ctt acc agt gga gca tac aaa cct gca gca gat 979  
 Ala Ala Leu Ala Ser Leu Thr Ser Gly Ala Tyr Lys Pro Ala Ala Asp  
 280 285 290

gaa aaa gtg gca gtc att gtg tgc gga gcg aac act gac ctc aca aca  
 1027  
 Glu Lys Val Ala Val Ile Val Cys Gly Ala Asn Thr Asp Leu Thr Thr  
 295 300 305

ctg tgatgtgatt tcaaacgata aca  
 1053  
 Leu  
 310

&lt;210&gt; 140

This Page Blank (copy)

&lt;211&gt; 310

&lt;212&gt; PRT

&lt;213&gt; Corynebacterium glutamicum

&lt;400&gt; 140

```

Met Leu Thr Leu Asn Asp Val Ile Thr Ala Gln Gln Arg Thr Ala Pro
 1           5           10           15

His Val Arg Arg Thr Pro Leu Phe Glu Ala Asp Pro Ile Asp Gly Thr
          20           25           30

Gln Ile Trp Ile Lys Ala Glu Phe Leu Gln Lys Cys Gly Val Phe Lys
          35           40           45

Thr Arg Gly Ala Phe Asn Arg Gln Leu Ala Ala Ser Glu Asn Gly Leu
          50           55           60

Leu Asp Pro Thr Val Gly Ile Val Ala Ala Ser Gly Gly Asn Ala Gly
 65           70           75           80

Leu Ala Asn Ala Phe Ala Ala Ala Ser Leu Ser Val Pro Ala Thr Val
          85           90           95

Leu Val Pro Glu Thr Ala Pro Gln Val Lys Val Asp Arg Leu Lys Gln
          100          105          110

Tyr Gly Ala Thr Val Gln Gln Ile Gly Ser Glu Tyr Ala Glu Ala Phe
          115          120          125

Glu Ala Ala Gln Thr Phe Glu Ser Glu Thr Gly Ala Leu Phe Cys His
          130          135          140

Ala Tyr Asp Gln Pro Asp Ile Ala Ala Gly Ala Gly Val Ile Gly Leu
          145          150          155          160

Glu Ile Val Glu Asp Leu Pro Asp Val Asp Thr Ile Val Val Ala Val
          165          170          175

Gly Gly Gly Gly Leu Tyr Ala Gly Ile Ala Ala Val Val Ala Ala His
          180          185          190

Asp Ile Lys Val Val Ala Val Glu Pro Ser Lys Ile Pro Thr Leu His
          195          200          205

Asn Ser Leu Ile Ala Gly Gln Pro Val Asp Val Asn Val Ser Gly Ile
          210          215          220

Ala Ala Asp Ser Leu Gly Ala Arg Gln Ile Gly Arg Glu Ala Phe Asp
          225          230          235          240

Ile Ala Thr Ala His Pro Pro Ile Gly Val Leu Val Asp Asp Glu Ala
          245          250          255

Ile Ile Ala Ala Arg Arg His Leu Trp Asp Asn Tyr Arg Ile Pro Ala
          260          265          270

Glu His Gly Ala Ala Ala Ala Leu Ala Ser Leu Thr Ser Gly Ala Tyr
          275          280          285

Lys Pro Ala Ala Asp Glu Lys Val Ala Val Ile Val Cys Gly Ala Asn
          290          295          300

```

Page Blank (uspto)

ist unsere Segel

216

**This Page Blank (uspto)**



tcc caa tta ctc gca tat ggt cgc gat ttt gcg gag gtc atg aag gat 643  
 Ser Gln Leu Leu Ala Tyr Gly Arg Asp Phe Ala Glu Val Met Lys Asp  
 170 175 180

aat gag cgc tta atc cac ggg gat ctt ggc aca gtg gat gcc cat ttg 691  
 Asn Glu Arg Leu Ile His Gly Asp Leu Gly Thr Val Asp Ala His Leu  
 185 190 195

gat cga gtg tgg cag att atg cag gag tgc gtg gca caa ggc atc gca 739  
 Asp Arg Val Trp Gln Ile Met Gln Glu Cys Val Ala Gln Gly Ile Ala  
 200 205 210

acg ccg ggg att tta ccg ggt ggg ttg aat gtg caa cgt cgg gcg ccg 787  
 Thr Pro Gly Ile Leu Pro Gly Gly Leu Asn Val Gln Arg Arg Ala Pro  
 215 220 225

cag gta cac gcg ctg att agc aac ggg gat acg tgt gag ctg ggt gct 835  
 Gln Val His Ala Leu Ile Ser Asn Gly Asp Thr Cys Glu Leu Gly Ala  
 230 235 240 245

gat ctt gat gct gtg gag tgg gtg aat ctg tac gcc ttg gcg gtg aat 883  
 Asp Leu Asp Ala Val Glu Trp Val Asn Leu Tyr Ala Leu Ala Val Asn  
 250 255 260

gaa gaa aac gcc gct ggt ggt cgt gtg gtt act gct ccg act aat ggt 931  
 Glu Glu Asn Ala Ala Gly Gly Arg Val Val Thr Ala Pro Thr Asn Gly  
 265 270 275

gct gcg ggg att att ccg gcg gtg atg cac tat gcg cgg gat ttt ttg 979  
 Ala Ala Gly Ile Ile Pro Ala Val Met His Tyr Ala Arg Asp Phe Leu  
 280 285 290

aca ggt ttt ggg gcg gag cag gcg cgg acg ttt ttg tat acc gcg ggt  
 1027  
 Thr Gly Phe Gly Ala Glu Gln Ala Arg Thr Phe Leu Tyr Thr Ala Gly  
 295 300 305

gcg gtg ggc atc atc att aag gaa aat gcc tcg atc tct ggc gcg gag  
 1075  
 Ala Val Gly Ile Ile Ile Lys Glu Asn Ala Ser Ile Ser Gly Ala Glu  
 310 315 320 325

gtg ggg tgt cag ggt gag gtt ggt tca gcg tcc gcg atg gcg gct gcc  
 1123  
 Val Gly Cys Gln Gly Glu Val Gly Ser Ala Ser Ala Met Ala Ala Ala  
 330 335 340

ggg ttg tgt gca gtc tta ggt ggt tct ccg caa cag gtg gaa aac gcc  
 1171  
 Gly Leu Cys Ala Val Leu Gly Gly Ser Pro Gln Gln Val Glu Asn Ala  
 345 350 355

gcg gag att gcg ttg gag cac aat ttg gga ttg acg tgc gat ccg gtg  
 1219  
 Ala Glu Ile Ala Leu Glu His Asn Leu Gly Leu Thr Cys Asp Pro Val  
 360 365 370

ggc ggg tta gtg cag att ccg tgt att gaa cgc aac gct att gct gcc  
 1267  
 Gly Gly Leu Val Gln Ile Pro Cys Ile Glu Arg Asn Ala Ile Ala Ala

Page Blank (uspto)

375  
 atg aag tcc atc aat gcg gca agg ctt gcc cgg att ggt gat ggc aac  
 1315  
 Met Lys Ser Ile Asn Ala Ala Arg Leu Ala Arg Ile Gly Asp Gly Asn  
 390 395 400 405  
 aat cgc gtg agt ttg gat gat gtg gtg gtc acg atg gct gcc acc ggc  
 1363  
 Asn Arg Val Ser Leu Asp Asp Val Val Val Thr Met Ala Ala Thr Gly  
 410 415 420  
 cgg gac atg ctg acc aaa tat aag gaa acg tcc ctt ggt ggt ttg gca  
 1411  
 Arg Asp Met Leu Thr Lys Tyr Lys Glu Thr Ser Leu Gly Gly Leu Ala  
 425 430 435  
 acc acc ttg ggc ttc ccg gtg tcg atg acg gag tgt tagcgggtacg  
 1457  
 Thr Thr Leu Gly Phe Pro Val Ser Met Thr Glu Cys  
 440 445  
 gctttaacac ggc  
 1470  
  
 <210> 142  
 <211> 449  
 <212> PRT  
 <213> Corynebacterium glutamicum  
  
 <400> 142  
 Met Ala Ile Ser Val Val Asp Leu Phe Ser Ile Gly Ile Gly Pro Ser  
 1 5 10 15  
 Ser Ser His Thr Val Gly Pro Met Arg Ala Ala Leu Thr Tyr Ile Ser  
 20 25 30  
 Glu Phe Pro Ser Ser His Val Asp Ile Thr Leu His Gly Ser Leu Ala  
 35 40 45  
 Ala Thr Gly Lys Gly His Cys Thr Asp Arg Ala Val Leu Leu Gly Leu  
 50 55 60  
 Val Gly Trp Glu Pro Thr Ile Val Pro Ile Asp Ala Ala Pro Ser Pro  
 65 70 75 80  
 Gly Ala Pro Ile Pro Ala Lys Gly Ser Val Asn Gly Pro Lys Gly Thr  
 85 90 95  
 Val Ser Tyr Ser Leu Thr Phe Asp Pro His Pro Leu Pro Glu His Pro  
 100 105 110  
 Asn Ala Val Thr Phe Lys Gly Ser Thr Thr Arg Thr Tyr Leu Ser Val  
 115 120 125  
 Gly Gly Gly Phe Ile Met Thr Leu Glu Asp Phe Arg Lys Leu Asp Asp  
 130 135 140  
 Ile Gly Ser Gly Val Ser Thr Ile His Pro Glu Ala Glu Val Pro Cys  
 145 150 155 160

This Page Blank (uspto)

Pro Phe Gln Lys Ser Ser Gln Leu Leu Ala Tyr Gly Arg Asp Phe Ala  
165 170 175

Glu Val Met Lys Asp Asn Glu Arg Leu Ile His Gly Asp Leu Gly Thr  
180 185 190

Val Asp Ala His Leu Asp Arg Val Trp Gln Ile Met Gln Glu Cys Val  
195 200 205

Ala Gln Gly Ile Ala Thr Pro Gly Ile Leu Pro Gly Gly Leu Asn Val  
210 215 220

Gln Arg Arg Ala Pro Gln Val His Ala Leu Ile Ser Asn Gly Asp Thr  
225 230 235 240

Cys Glu Leu Gly Ala Asp Leu Asp Ala Val Glu Trp Val Asn Leu Tyr  
245 250 255

Ala Leu Ala Val Asn Glu Glu Asn Ala Ala Gly Gly Arg Val Val Thr  
260 265 270

Ala Pro Thr Asn Gly Ala Ala Gly Ile Ile Pro Ala Val Met His Tyr  
275 280 285

Ala Arg Asp Phe Leu Thr Gly Phe Gly Ala Glu Gln Ala Arg Thr Phe  
290 295 300

Leu Tyr Thr Ala Gly Ala Val Gly Ile Ile Ile Lys Glu Asn Ala Ser  
305 310 315 320

Ile Ser Gly Ala Glu Val Gly Cys Gln Gly Glu Val Gly Ser Ala Ser  
325 330 335

Ala Met Ala Ala Ala Gly Leu Cys Ala Val Leu Gly Gly Ser Pro Gln  
340 345 350

Gln Val Glu Asn Ala Ala Glu Ile Ala Leu Glu His Asn Leu Gly Leu  
355 360 365

Thr Cys Asp Pro Val Gly Gly Leu Val Gln Ile Pro Cys Ile Glu Arg  
370 375 380

Asn Ala Ile Ala Ala Met Lys Ser Ile Asn Ala Ala Arg Leu Ala Arg  
385 390 395 400

Ile Gly Asp Gly Asn Asn Arg Val Ser Leu Asp Asp Val Val Val Thr  
405 410 415

Met Ala Ala Thr Gly Arg Asp Met Leu Thr Lys Tyr Lys Glu Thr Ser  
420 425 430

Leu Gly Gly Leu Ala Thr Thr Leu Gly Phe Pro Val Ser Met Thr Glu  
435 440 445

Cys

<210> 143  
<211> 1425

**This Page Blank (uspio)**